

歴史研究に開かれた学籍データベース構築の課題

——旧制兵庫県立神戸第一中学校を事例として——

加藤 善子 保田 その

キーワード：データベース作成 個票データ 近代日本 神戸一中

1. 問題意識

1-1.はじめに

本稿は、歴史研究に供するための個票データベース作成の方法論を確立するべく、その課題を整理するものである。この課題は、我々が共同研究の一環として実際に学籍データベースを作成した際に直面した、具体的な問題を解決するための必要性から生まれたもので、これまでに多くの研究者が直面してきた課題でもある。

本稿の議論は、1896（明治26）年から1944（昭和19）年までの旧制兵庫県立神戸第一中学校（以下、「神戸一中」と略する）の学籍データと、そのデータによって得られた情報や知見をベースにしている¹。個人情報保護を背景に学籍簿や同窓会名簿等の利用が困難になる中で、このような個票データの希少性はますます高まることが予想される上に、科学研究費の助成を受けた研究として、データベースへのアクセスを多くの研究者に保証できるように貢献したいと考えた。そのためには、新たにこのデータベースに対面する研究者にとって、物理的にだけでなく、操作性においてもアクセスが容易なものでなくてはならず、従って、さらなる分析に開かれたデータベースを作成するという作業自体が、本研究の大きな目的の一つとなった。

この研究は教育社会学における歴史研究から出発しているが、この問題意識は、教育社会学および隣接領域において、学校を媒介とした社会移動研究がある程度蓄積されたからこそ生まれたものだといえる。そして、個票データの分析は、社会移動研究を大きく進めるものであった。既存の統計資料からは個別の変数ごとの単純集計しか手に入らず、研究者が自らデータ加工を施す余地はほとんどなかった状態から、変数を組み合わせてその関連を自在に分析できるようになったのだ（河野, 2001, pp.169-170）。

教育社会学における歴史研究は（そして一般的な歴史研究も同様に）、入手できるデータに制限があったため、それぞれの資料を入手した研究者がその資料の特性によりカテゴリー化を行い、それぞれ分析した研究成果が断片的に集積されるという状態になっている。それらの断片的な研究成果は、データや資料がもつ限定的な性質のた

めに、異なる仮説が出現した場合に再分析が不可能なものがあったり、あるいはカテゴリーの不一致などにより、単純比較すら不可能であったりする。これまでの研究の蓄積がうまくリンクすれば、あるいは、リンクが可能で、加工自在なデータがあれば、研究の飛躍的發展が期待できる。

教育社会学における歴史研究との接続や関係性を意識して、これまで異なるレベルで行われてきた研究の蓄積と対話を可能にするようなデータベースを構築するために、まずは取り組むべき課題を整理し（本節）、隣接領域において模索されてきたデータベース作成に関する方法論をレビューして（第2節）、そして、具体レベルでの問題と解決策を提示する（第3節）。さらに、個人情報保護や個人情報の利用ガイドラインなど法令の変遷を、補論として文末にまとめることとした。今後、新しい分析視角や仮説が登場した際に、さらなる分析に開かれたものとして本研究のデータベースが利用されるようになることも想定に入れて、考察したい。

1-2. 先行研究と課題の設定

1-2-1. 教育の歴史社会学における成果

この研究分野は、近代化に伴う、学校を経由した社会移動と社会の構造変化を明らかにして着実に研究成果を蓄積してきた。同時に、資料によって研究対象や範囲が限定される性質から免れない歴史研究の蓄積は、「習作群」として積み上げられてきたという。広田（1995）は、これまでの歴史の「〈習作〉群」の問題として、その権威主義的な性格と視点の単純さを指摘している（広田, 1995, p.31）。特に後者については「伝統／近代という単純な二分法に陥っている」ことが指摘されている。広田が指摘するように、これまでの歴史の「〈習作〉群」は近代に特有なものを見ようとし、近代以前のものも近代の中の細かい変容も見ようとしないために「直線的な趨勢」が再確認されることにもなる（広田, 1995, pp.33-34）。

教育を媒介とした社会移動研究においても、確かに、この直線的な趨勢は確認することができる。「伝統セクター」から「近代セクター」へ、「農村」「地方」から「都市」へ、「旧中間層」から「新中間層」への、教育を媒介にした人口の移動という趨勢を追って、日本の近代化の過程を記述することに多くのエネルギーが費やされてきた。これらの個別の研究の蓄積を俯瞰し相互の齟齬を検討してモデル構築を導く点で発展が期待されるが（広田, 1995, p.33）、そこで立ちはだかる限界や障害は、資料に依存する、あるいは研究者の個別の関心に依存するカテゴリー化の多様性、そこへ至る思考過程と手続きの記録の欠如である。

各種統計書を使用した研究は非常に多い。対象地域や学校を限定して、学校や地域の一次的資料やその他の情報を地域の公式統計情報に照合する作業を行いながら、学校化や近代化の過程を追う研究は、天野（1991）、吉田（1985）、土方（1994; 2002）、寺崎・吉田（2000）、湯田（2010）、武石（2012）など着実に行われてきている。全国

の公式統計情報を集め全国レベルでの進学機会を分析したのが菊地（2003）であり、この研究により、各地域の相対的な位置関係が把握できるようになった。

ケースを限定してその内部集団の相互作用や関係構築・変容のダイナミクスを把握しようとするならば、比較の手段としての集団のカテゴリーを統一しなければならない。統計書や学籍簿で使用されているカテゴリーをそのまま分析に使用している研究もあれば（望月, 1987 ほか）、それらを再カテゴリー化して元のカテゴリーと併用して分析している研究もある（吉田, 1985; 保田・薄葉・竹内, 1999; 寺崎・吉田, 2000 ほか）。現時点ではそれほど重要なカテゴリーとみなされていないにせよ、宗教家が新中間層に分類されている研究もあれば（吉田, 1985）、専門職として分類されている研究もあり（保田・薄葉・竹内, 1999）、また鶴岡中学校の個票を使った一連の研究では新中間層でも旧中間層でもなく「その他」に分類されている（鈴木, 2001）一方で、第一回国勢調査では「公務・自由業」に分類されている（鈴木, 2001, p.195）。構造移動の影響があるにせよ、出身階層を表すときの「新中間層」の下位区分と、到達階層を表すときの「新中間層」の下位区分は同一でない場合もあり、比較参照する場合に注意を要する。

1-2-2. 神戸一中データが提起する問題

神戸一中の学籍データの提供を受け、データベースの作成を進めながら分析を試みたものの、このデータを使った論文を発表していくには困難が伴った。というのも、これまでの研究によって確認されてきたような、直線的な近代化の趨勢にあてはめて解釈することのできないデータだったからである。

まず、神戸一中入学者父兄の職業構成を集計すると、学校利用に消極的な「遅れた旧中間層」と従来の知見ではみなされてきた商工業層（天野編, 1991）が30%を超えており、会社員が40%にのぼる一方、官公吏はわずか9%、農業も1割に満たないことが明らかになった（加藤, 2011, p.180）。官公吏と農業の割合は時代を下るにつれて更に減っていく。年代ごとに区切って保証人の職業構成の趨勢を追うと、神戸一中を利用したのは産業地区の商業層（「旧中間層」に分類される）が最初であり、その後に住宅地域から会社員層（「新中間層」に分類される）が参加する傾向が見られ、この2種類の進学パターンが明治後期から昭和初期まで持続したのである（加藤, 2011, p.187）。

職業移動パターンを分析したところ、神戸一中のケースでは、「実業セクター」内で再生産が起こる傾向が見られる。商業層と会社員の間で再生産がおこり、その移動は双方向的であることが確認された。「新中間層」「旧中間層」という区別よりも、「実業セクター」「非実業セクター」という経済史的カテゴリーを用いる方が、神戸では職業構成や経済構造と学校利用戦略との関係をより理解しやすい（井上, 2006）²。この知見は、「旧中間層」から「新中間層」への移動は一方的であり、「新中間層」に移動した後に、「新中間層」内部で再生産が起こるという、鶴岡を含む先行研究の多くが確認してきた現象とは一線を画するものである。

その他にも、このフィールドが「都市」であることによる分析視角の転換が必要になる。神戸市（およびその周辺）の出身者にとっての進学に伴う社会移動は、「都市への」移動ではなく「都市からの」移動、もしくは「都市内部での」移動として読みかえる必要が出てくる。また、人口の急激な増加や市域の拡張、隣接区域の扱い方（神戸には大阪や京都などの大都市からも通学可能である）など、これまでの歴史研究の対象が主に農村地域であったことに由来する直線的近代化モデルを相対化する示唆に富むデータであることは間違いない。

とはいえ、新しいデータを使って新しい仮説を打ち立てるだけでは、先行研究との関連性、相似点、相違点を規定する要因を抽出するための実証的な手続きをとることができない。その意味で、様々なカテゴリー化の作業に耐え、各種統計との対話を可能にするデータが必要なのである。この作業のモデルの一つとなるのが、粒来・佐藤（1995）である。65年SSMデータの再々コード化をしているが、SSMデータが再加工に耐えるデータであることをはじめ、そこで試みられた再コード化と合成変数作成の手続きが、分析に先んじて正確に記述されている。データの代表性が確保されていることから検定が可能であり、確実に研究成果として積み重ねられていく。

この遡及性を確保することが研究発展のために不可欠な方法として確立されるべきだと我々研究グループは痛感させられ、遡及性の確保が拡張性の実現や公開性の問題と同根であることに考えが至るようになったのである。

1-2-3.直接の先行研究：鶴岡研

その意味で、本研究は、広田を代表とする研究グループの成果（2001）を何よりも第一に依拠すべき先行研究として特別の敬意を払い、参照している。鶴岡中学の学籍情報をデータベース化し分析したこのグループの研究は一次資料（個票）を利用した最初のもののひとつである³。個票データに基づき、中学内部の選抜過程を詳細に分析し、経済状況や族籍による進学パターンをより精緻に分析して新しい知見を多く出したが、成果はそれだけではない。むしろ本研究が目にするのは、彼らのデータとの対話と対峙の経験であり、データベース化に際しての様々な課題と問題解決の過程である。

河野が指摘しているように、個票データによって「既存の統計資料からはうかがい知ることのできなかつた諸側面を明らかにしていくことが可能」となり、データベースを操作すれば、一般の社会調査データのように解析することができる（河野, 2001, pp.169-170）。その資料が持つ「固有の問題」を十分理解した上で作成されたデータは、「データ加工の余地の大きさ」と、その「拡張性」という二つの性質を持つ（河野, 2001, pp.178-179）。第一の点については、そのデータの正確さと範囲を担保するため、既存統計との照合作用が不可欠であり、その作業によってデータを「開かれた」ものにすることができる。この点は第二の性質に関係するが、他の中等教育諸機関との「ヨコ」の比較や、初等教育や高等教育との「タテ」の接続を可能にする（河野, 2001, p.179）。

本研究は、河野（2001）や鈴木（2001）によって指摘されたこの立場を共有する。

データベース作成にあたっては、第一に、「データ加工の余地の大きさ」の確保、つまり「遡及性」を確保して、様々な問題関心に応えるデータ加工を可能にするデータベースを作成しなければならない。その上で、第二の点である「拡張性」を確保する。たとえば、鶴岡中学と神戸一中のデータを比較したり、教育社会学における歴史研究の豊かな研究成果や既存統計との「相互参照的な関係」(河野, 2001, p.178)を築いたりするための準備が必要になる。

鶴岡グループの研究成果報告書では、上に述べた「カテゴリー化の多様性」や「そこへ至るまでの思考過程と手続きの両方の記録の欠如」に対して細心の注意がはらわれる。特に親の職業のカテゴリー化に際しては、「新中間層」「旧中間層」の対比という分析方針と、これまでの研究との一貫性の確保という目的とあわせて、なぜ一定の具体的なカテゴリー化の手続きをとるかを説明し、その過程を問題点および留保点とともに詳細に記述している(鈴木, 2001)。

1-3.課題：データの遡及性・拡張性・公開性

データベース作成者とは別の興味関心をもつ研究者が、カテゴリーや手法を変更して行う分析に耐えるデータベースを構築するには、作業方針の決定、入力、修正、加工、さらに何らかの仮説に従った「変数」としてのカテゴリー化と分析、という一連の過程の全てにおいて、常に遡及性の確保に留意することが必要となる。これにより、今後新たに作られるものを含めた他のデータベースに基づいて行われる研究とも比較が可能になるという拡張性が生まれる。

そのために必要になる作業は、データベース内のデータの出所の明記、処理・分析過程の記録を残すことである。資料の記載内容はもちろん、何らかの操作を行った場合にはその過程やコーディングのルールをすべて可能な限りもれなく記録することが求められる。変数化のプロセスを可視的な状態で残し、将来の再分析などにおいて遡及可能なものにすることが、データベースの利用価値を決定する。次節で詳細に記述するが、複数ある神戸一中の資料から必要な情報を取りだし、時代や様式の違いに由来するカテゴリーやデータの揺らぎを統一し、それらの情報を一つのデータベースに統合する過程そのものの情報と記録が、研究成果であり財産になる。

さらに、現実にこのデータベースを公開して研究グループ外の利用に供する際には、その公開可能性が大きな課題となる。これについては個人情報保護、資料の管理責任などの問題を踏まえて検討することが必要となる(この点については、文末の「補論」にて触れる)。

このような視点と作業過程の記録という営みは、従来、個票データを用いた研究においてはほとんど取り入れられることがなかった。分析結果が論文や著作となった後に、データベースを公開するという発想がこれまでほとんどなかったことや(研究者や研究グループの財産であるという考え方が支配的であった)、後に述べる個人情報保護やデータの提供元との信頼関係といった問題によって、不可能だと考えられてきたのだろう。今後のデータベース作成にあたっては、このような観点からの蓄積が比較

の進みつつある分野における知見を援用しながら、方法論を蓄積していくことが課題になる。

2. 方法論のレビュー

2-1.他領域の方法論

本研究が援用することが可能な分野として考えられるのは、①社会調査データの二次分析、②作成された当初の目的が研究の目的と一致しない研究分野における資(史)料の整備・加工の方法論(歴史人口学や識字統計における宗門改帳や宗門人別帳などのデータベースの利用)、③文書館的施設における文書資(史)料(archival documents)の目録作成や公開、という三つの分野である。これらの分野においても、データベースの遡及性、拡張性、公開可能性、といった観点から、本格的に研究の蓄積が進み始めたのは最近である。

第一の分野である社会調査データの二次分析に関しては、佐藤・石田・池田が、日本において社会調査データの二次分析を用いた研究が少ないことを指摘し、その背景として、自分でデータを集める一次分析が高く評価されること、また収集されたデータが公開されることが非常に少ないことを挙げている(佐藤・石田・池田, 2000, p.1)。とはいえ、1998年には東京大学社会科学研究所附属日本社会研究情報センターが、SSJデータアーカイブ(Social Science Japan Data Archive)を設立したことにより、徐々に多くの調査データベースが蓄積され、利用に供されるようになった。SSJデータアーカイブのウェブサイト(http://jww.iss.u-tokyo.ac.jp/overview/2_2.html, 2017年1月10日取得)によると、「調査主体が集計結果を出し、分析して発表すれば、もともなったデータは放置されてしまい」、「他の研究者が追試したり、別の観点から分析しなおしたりするということは不可能」であるため、分析結果の客観性が担保されず、かつ「通常かなりの費用を必要とする社会調査の結果が社会的に共通されない」ことになる。「とくに、科学研究費のような公的資金を用いた調査の場合には、公的資金の有効利用という観点からも問題」であるとして、調査データベースの公開が進んだ。

第二の分野として挙げた、そもそも作成当初の目的や価値から離れたところでその資(史)料から情報を引き出そうとする研究分野に蓄積されている方法論は、歴史人口学における宗門改帳や宗門人別帳などのデータベース化作業などが本研究と問題関心を共有している。ある特定のテーマと資料の使用にあたっての代替変数の設定過程や他の研究データとの比較には、その記録が詳細に記されていて参考になる。

宗門改帳のデータベース化と分析の成果をまとめた速水は、研究行程をふりかえって、歴史学・人口学・社会学などの複数ディシプリンの習得が必要だと途中で気付いたと言い、「本格的な作業はこれから」と述べていたが(速水, 2002, p.16)、歴史人口学の分野においては、従来、個々の村などを単位として収集、分析されてきた長期ミクロデータを、国際比較の可能なデータベースとして構成するということが1990年代末より行われつつあった。ここに様々な課題があることを、黒須(2008)は“Eurasia

Project”⁴に参加した経験に基づいて指摘している。

黒須によると、「日本の歴史人口学者が利用している宗門改帳や人別改帳は、基本的に毎年の人口静態（世帯ごとに存在するメンバー）や持高を含む世帯情報とともに人口動態（出生，死亡，結婚，移動など）のイベントが記載」され、「詳細かつ複雑な情報が含まれている」ものであるが、「入力されたデータが数多くありながら，デジタルデータ活用が遅々として進まない」という（p.51）。その要因として、「(1)データの情報量と史料のプロセスの仕方、(2)データの持つプライバシーや利用権の問題、(3)長期マイクロレベルデータの扱い方の3点」が存在する。そして特に（1）に関しては、「速水プロジェクトで行われてきたデータ構築までの道のりは長く複雑で、長い間『奉公』して理解しなければ利用できないという性質」が、「ある一つの村に専念して研究を進める」「オリジナル史料をそのままに利用する」という研究方法の特徴によって問題にならなかったことを指摘している（黒須，2008，p.52）。アーカイブズ（永久保存記録や歴史資料）の保存や活用に必要な情報の種類とその管理についての実績に学びながら「今後地道にそれらの情報を整理保存していかななくてはならない」として、ICPSR（2007）や村越（2007）の先行研究に言及している（黒須，2008，p. 53）。

さらに第三の分野の主体である公文書館、資料館といったアーカイブズの施設においても、従来は歴史研究者や郷土史家が独自の観点からのみ目録を作成していたが、近年は、文書目録のデータベースをインターネット上で公開することが多くなることに伴って、目録記述の方法を標準化する試みが進められた。同時に、図書に比べて形式に統一性のない文書資料の目録を如何に記述するかという方法論が蓄積されつつある（鈴江，2002）。特に近年では、五島（2007）、村越（2007）らが、資料管理における作成記録情報の記述の重要性を指摘しており、歴史人口学においてもその知見を援用する試みが始まっている。

また、このようなアーカイブズの施設においては所蔵資料を閲覧に供することが主要な業務であり、特に近年では、個人情報の保護に配慮しながら、閲覧、公開の基準を定めている（個人情報保護については、文末の「補論」で触れる）。

以上の分野は、我々の分野である教育社会学とは性質がかなり異なる面もあるが、問題意識は共通しており、これらの分野における蓄積を取り入れながら方法を模索することができると思われる。

2-2. データ入力作業における「遡及性」の確保

次に、以上のような他分野の知見を取り入れながら、遡及性、拡張性が確保されたデータベースを作成するための方法論を探っていくことにしよう。まずは、データ処理の第一段階であるデータ入力作業について考えてみたい。

社会調査のうち定型的な質問によって構成された質問紙調査などのデータ入力は、特に選択式の設問の場合、外注や単純作業のアルバイト要員が機械的に行うことで十分である。むしろ、質問紙を作成する段階で、機械的な入力作業が可能なものを作成するのが通例である。

他方、ここで扱う『学籍簿』や『参考簿』のように、本来は、学校が生徒の背景や進路を把握するために作成されたものに関しては、当然ながら、機械的な入力に適した形につくられているわけではないので、質問紙調査のような機械的な入力では不十分な場合がある。

このような判断を要する入力作業の方法は確立されているとは言い難いが、先に述べたアーカイブズにおける史料管理の分野における「目録記述」に関する方法論を部分的に援用することができるだろう。史料管理の分野も、我が国においては決して蓄積のある分野とはいえないが、それでもこの十数年間で整備が進んでいる。

定型的なアンケート調査における調査票の入力作業と、学籍簿に代表されるような業務上の必要性に応じて作成された個票の入力作業との違いは、図書館における新刊書の目録作成作業と、文書館・資料館における文書資料の目録作成作業との違いに相当すると考えられる。近年では、文書館、資料館における文書の公開が進む中で、目録作成のマニュアルを確立することが急務となり、マニュアルの例やいわばメタ方法論に当たる議論が行われている。

このうち、鈴江は、「客観的な表示によって記述しうるものと、目録作成者の判断が作用して記述するものとは区別されるべきである」ということを強く主張し、そのための方法論、たとえば、補足情報を記入するための括弧類の使用法の確立などを提案している（鈴江, 2002, pp.340-342）。

3. 実際に生じた問題とその解決

ここからは入力作業において生じた様々な問題について述べていく。今回使用した、神戸一中の学籍データを構成する資料には、『学籍簿』（明治29年～昭和19年、入学年度を単位として編纂）、『（主任）参考簿』（明治43年～昭和15年、卒業年度を単位として編纂）、『半途退学生徒主任参考簿』、『同窓会名簿』（『同窓会報』の巻末付録、昭和16年・昭和18年）、『同窓会報』（大正15年～昭和18年・昭和23年）、『交友会誌』（明治33年～昭和23年）がある⁵。

3-1-1. 記載内容の修正

まずは、『学籍簿』や『参考簿』の記載事項が修正されている場合について考える。質問紙調査であれば、修正前の記載は回答者の勘違いなどによるものであり、修正後の記載を採用するのが一般的であるだろう。一方、個票データの場合、抹消や修正の扱い方は一様ではない。

ここでは、例として、【正保証人職業】（分析においてはカテゴリー化した上で【親の職業】の変数として用いた）の項目欄で、記載が修正されている場合を考えてみよう。この欄における【正保証人職業】の修正例としては、主に三つのパターンがあった。

パターン①：「〇〇株式会社勤務」を線で抹消し、横に「無職」と加筆してある。加筆修正は、元の記載とは別の筆記用具で行われている。その横にある「正保証人氏名」には男性の名前が、「正保証人との関係」には「父」と書かれており、これらの欄には何の修正もない。

パターン②：「会社員」を線で抹消し、横に「無職」と加筆。加筆修正は、元の記載とは別の筆記用具で行われている。その横にある「正保証人氏名」には当初男性の名前が記載されていたが、それが抹消されて女性の名前が加筆されており、「正保証人との関係」欄に当初「父」と記載されていたものが抹消されて「母（父死亡）」と加筆されている。

パターン③：「庶業」を線で抹消し、横に「小学校教員」と加筆。加筆修正は本の記載と同じ筆記用具、筆跡で行われている。その横にある「正保証人氏名」、「正保証人との関係」欄には何の修正もない。

これら三つのパターンは、それぞれ意味するところが大きく異なる。また、分析の対象とする「正保証人職業」を、入学時点のものとするか、卒業（退学）時点のものとするかでも異なる。基本的には「正保証人職業」は【親の職業】についての情報に代わるものとして、さらにいえば、本人の出身階層や家庭背景の変数として用いるものである。

まずパターン①は、おそらく生徒本人の入学時には「〇〇株式会社」に勤務していた正保証人である父親が、生徒の在学中におそらくは定年などで退職したと考えられる。パターン②は、父親の職業は「会社員」であるが、生徒本人の在学中に、父の死亡という不慮の出来事があり、「無職」の母が代わりに保証人になったと考えられる。いずれのケースも、本人の出身階層の指標として用いる「正保証人職業」として考えれば、変数としては修正前の記述を使うべきだ。

一方、パターン③の場合はこれと異なる。これはおそらくはかつて「農工商以外のその他の職業」を意味する者として頻りに用いられていた職業分類である「庶業」（鈴木, 2001, p.195）という用語を一旦は用いたが、考えなおしたり、誰かの助言を受けたりして、具体的な職業名である「小学校教師」を使った方がよいと判断して書き直したと考えられる。よって、この場合、分析には修正後の「小学校教師」を使うという判断が正しいだろう。

問題は、データベースへの入力時に、入力者がこのような判断を行うことは不可能であることである。さらに仮にパターン①やパターン②のようなケースであることが確実であっても、修正後の職業記載を無視するということがあってはならない。他の分析において「正保証人職業」を別の指標として分析する時に、修正後の職業を用いることのできる可能性を残す必要がある。

3-1-2.記載欄・項目名の変動

たとえば、【親の職業】という変数を例にとって考えてみよう。これについては『学籍簿』の一部（明治36年入学、大正12年入学が中心）と、『参考簿』（全範囲、明治44年11回卒業～昭和2年卒業）に記載されている「正保証人職業」を用いることができる。

『参考簿』における「正保証人」と本人との関係は、「父」（同義の記載を含む）が全体の約80%を占めていること、「正保証人職業」の欄の空欄は全体の2%弱に過ぎないことを考えると、この記載内容を『親の職業』として扱うのが適当である。だが、一部には、父親が健在であっても正保証人が父親ではなく、他人であるというケースもある。そのような場合、【正保証人職業】の欄にはその他人の職業が記載されるが、【備考】欄に「父は〇〇会社勤務」といった形で父の職業が書かれたり、【家族】欄に父の氏名を書いた横に「(△△商)」のような形で父の職業が書かれたりする場合がある。

この場合、【親の職業】という変数を、大部分は「正保証人職業」の記載事項を用い、そこに情報がない場合、あるいは正保証人が親以外の場合で別の欄（備考欄、家族欄など）に親の職業が書かれている場合には、それらを用いる方法が現実的である。しかし、だからといって、最初から【親の職業】欄に、あるケースでは【正保証人職業】の情報を入力し、別のケースでは【備考】の情報を入力するというようにしていたのでは、入力者の判断にばらつきがでることが考えられる。

たとえば、入力者は、もし、単純に「正保証人の職業を入力せよ」という指示のみを受けていたならば、備考欄や家族欄を無視して、【正保証人職業】欄の情報のみを入力することになるだろう。「親の職業に関する情報が必要なので、【正保証人職業】欄を入力せよ」という指示を受けていたとすれば、入力者は、場合によっては、【正保証人職業】欄に書かれている他人の職業は無視して、【備考】欄や【家族】欄にある父の職業を入力するかもしれない。

前者の方法であれば、そのサンプルに関しては【父の職業】に関する情報が抜け落ちてしまう。後者の方法であれば、その入力に、かなり恣意的な判断が入ることになる。「正保証人が父以外であり、かつ備考欄や家族欄に父の職業が書かれていた場合には、【正保証人職業】欄の記載は無視して、そのかわりに父の職業を入力せよ」というのは複雑すぎる指示である。

3-2.入力作業における具体的な解決策

以上のような問題点を踏まえて、遡及性を確保するためには次のような注意が必要である。まず、個票の内容をそのまま写して入力する部分と、入力者もしくは分析者が判断を加えた部分を明確に区別することである（鈴江, 2002）。つまり、資料に含まれる記載事項と、分析に用いる変数とを区別して扱う、ということだ。

具体的な方法としては、入力者は記載内容をそのまま写して入力するが、入力になじまない記載が出た場合、その旨を【】内に記入して、併記することが挙げられる。たとえば、ある記載があつて、それを線で抹消して別の記載がある場合には、「A【後

に線で抹消して「B」と加筆あり】あるいは「B【脇に「A」と記載して線で抹消した形跡あり】」といった具合である。また、判読しにくい文字があった場合には、何らかの記号で代入し、【】内に「◇は@偏に△の右側のような字」というように加筆する。

【】内に記録する際は、入力者の判断を詳しく書くことを重視する。入力者が入力の仕方に迷ったとき、少なくとも【要チェック、迷った】という記述があれば、別の人がその個票に再び戻って確認することができる。【「〇〇」という記載が鉛筆で書かれており、その横に「△」とペンで書かれているので迷った】と書かれていれば、場合によっては、再び個票にあたらなくとも最終的な判断をどうするか決めることが可能である。だが、何の補足説明もなかったとすれば、当初の入力者の判断によってそのデータが決まってしまうことになり、後に誰もそれを疑うことができなくなってしまう。

後に、その記載内容を、分析用の変数として使用する場合には、別の入力欄を用いて入力する。素入力とカテゴリー化された変数の二者だけでなく、その間のプロセスに対しても別の欄（スプレッドシートの「列」）を用意することが重要だろう。

3-3.入力内容のカテゴリー化における事例と解決策

さらに、カテゴリー化を行う際には、その際の判断基準およびカテゴリー化した変数と、元の記載内容との対応表を作成し、再カテゴリー化やカテゴリーの見直しが行えるようにしておくことが重要であろう。

カテゴリー化の判断基準は決して一様ではなく、しかも、確実に正解だと言える判断基準はない。以下はいずれも、判断が出来ない場合である。

- ①異なるカテゴリーに属する複数の記載（一方が抹消されている場合も含む）がある場合
- ②いずれのカテゴリーに入れるべきかを判断しにくい場合
- ③カテゴリーの範囲が時期によって異なる場合

①にあたるものとしては、出身地域や正保証人職業などに、加筆修正、複数列举などがあり、異なるカテゴリーに属するものが記載されている場合である。たとえば、【正保証人職業】において、「会社員、元〇〇委員会代表」とある場合、「無職、元教員」とある場合、「地主、会社員、市会議員」と列举されている場合、どれをとるかは判断が難しいが、研究グループ内で基準を定めることが重要であろう。そしてさらに重要なのはその判断基準を、文章化したマニュアルのかたちで残し、共有することである。

②にあたるものは『同窓会名簿』に多い。同窓会名簿の職業欄には、多くの場合、所属する勤務先の名称が書かれている。【本人の職業】という変数は、勤務先をいずれかのカテゴリーに当てはめることによって作成するものであるが、当然、カテゴリー化にあたって判断の難しいケースが多く見られる。

たとえば、田中太郎氏（仮名）の勤務先欄が「住友商事株式会社」と書かれていたら、「会社員」というカテゴリに入れるだろう。同じ田中太郎氏の勤務先が「田中商店」であったとすれば、おそらくは個人商店を営んでいるものと考えて、「商業」「商工業」というカテゴリに入れることになるだろう。では、もし鈴木一郎氏（仮名）の勤務先が「鈴木商店」であったときにどうするか。鈴木商店が神戸を拠点とした大規模な商社であったことはある程度知られている。その知識をもって、我々はふつう、勤務先欄の「鈴木商店」を会社員というカテゴリに入れる。だが、親から引き継いだ個人商店である「鈴木商店」を営む卒業生がいないという保証はない。よって、最終的な方法としては、他のデータの傾向などから判断して（間違った判断である可能性があることも認めながら）、いずれかのカテゴリに入れるか、あるいは、そのようなカテゴリの作成をあきらめるしかない。

だが、ここでも大事なことは、このように判断したという記録を決して消さないことである。将来このデータを分析する人が、再びそのカテゴリを再考する機会を奪わないように万全を期すべきだ。「今回はこういう事情で乱暴な判断を行って分析をした」という記録を残しておくことが重要なのである。

③にあたるものとしては、住所に用いられている市町村が、合併によってその境界線を変化させているという問題点がある。たとえば神戸市は周辺町村の編入を伴う合併を繰り返しており、どの時点での「神戸市」を出身地の単位として分析するかは大きな課題である。

また、進学先の学校も時期によってその属するカテゴリが変化する。卒業者の多くが進学した神戸高等商業学校は昭和 4（1929）年に神戸商科大学に昇格した。このような場合に、昇格の前後を同じカテゴリに入れることは可能かどうかという問題点が生じてくる。そのカテゴリを用いることによって何を分析したいのか、仮説をあらかじめ明らかにした上で、その仮説を検証するために適切なカテゴリを設定する必要がある。

4. まとめにかえて

我々は、旧制中学の学籍データベースを作成するにあたって、上記に述べたような様々な困難に直面してきた。そして、この困難は、教育社会学の歴史分野にとどまらず、歴史人口学などの分野においてもほぼ共通するものであることが明らかになった。この困難さは、従来、社会学においても、歴史学や歴史人口学においても、データベース作成者、分析者、研究者がほぼ同一であり、また特定の研究対象にコミットしながらデータ作成と研究を並行して行う、という研究のスタイルが重視されるという伝統があったことにより、定型的な入力作業の場合を除いては、方法論の確立ということがほとんど求められてこなかったことによるところが大きい。

しかし、今後、オリジナル史料に戻らずデータベースを用いて研究を行うスタイルの研究が増加することは不可避であり、また必要でもある。なぜなら、研究とデータベース作成を一体化した形で特定の研究グループが排他的に行うという従来形では、

特定の村落、あるいは特定の学校などに関する個別研究とはなっても、社会移動一般の研究として普遍的なものにすることは難しく、ましてや国際比較などは非常に困難である。

特定の村落、特定の学校を対象とした個別研究を、社会学や歴史学の普遍的な研究の文脈に乗せるためには、個票データベースの作成において、その作成過程を逐一記録し、後の様々な利用可能性を開いておくこと、また、そのための技術論、たとえばアーカイブ資料における目録作成の技法などを取り入れることがますます重要になるだろう。

補論：データの公開可能性と個人情報の保護

個人情報を含むデータの扱いが難しくなった背景には、「個人情報の保護に関する法律」（2003年成立）がある。この法律の制定直後、多くの自治体が個人情報を含む文書を廃棄したり、従来閲覧に供していた資料の閲覧をとりやめたりすることが多く見られた。SSJ データアーカイブが公開を始めた当初は、個人情報保護に関する関心の高まりを受けて、プライバシーについても言及がなされていた。SSJ データアーカイブのウェブサイトでは、「プライバシーの保護」に関して、「回答者を特定できるような情報は公開し」ないこと、「利用者には、個々の回答者等が識別できる形式では発表しないことの制約を義務付ける」ことなどが対応策として提示されていた。「誤用の防止」に関しては、「調査内容や調査方法を十分理解して分析するよう指導」し、「調査ごとに『利用上の注意』を整理し、利用者に提供するとともに、調査報告書を収集し、閲覧できるように」することなどによって対応を行っているとの記載があった（<http://ssjda.iss.u-tokyo.ac.jp/about.html>, 2008年9月13日取得、2017年1月時点でこの記述は既がない）。

プライバシー保護という観点では、質問紙調査の多くが無記名で行われ、最初の分析の時点から回答者の氏名は特定できない場合が大部分であると考えられるのに対して、学籍簿を用いた研究においては、学籍簿に氏名が記入されており、データ入力やデータベース管理において氏名を用いざるをえない場合があるという問題点が残されており、これをどのようにクリアしていくかが、大きな課題として歴史研究者を悩ませた。

ただしこれに関しては、その第50条の除外規定で、「三 大学その他の学術研究を目的とする期間若しくは団体またはそれらに属する者 学術研究のように供する目的」の場合は適用除外となることが、また第三項では「第一項各号に掲げる個人情報取り扱い事業者」つまり学術目的に個人情報を利用するアーカイブズなどの機関は、「個人データの安全管理のために必要かつ適切な措置、個人情報の取り扱いに関する苦情の処理その他の個人情報の適正な取り扱いを確保するために必要な措置を自ら講じ、かつ、当該措置の内容を公表するよう努めなければならない」と定めている。

実際に、個人情報を含む歴史資料を扱う文書館は、個人情報保護法の成立に伴って

独自に閲覧制限をかける措置をとり、『独立行政法人国立公文書館利用規則』（平成13年4月2日規程第7号、最終改正平成18年6月20日）をその一例として参考にすることができる。他の文書館や資料館においても、国立公文書館のこの規程に準拠する形で独自の利用規程を定めたところが多く、少なくとも個人情報保護法の成立当初は説得力のある規程であったといえる。

この利用規則の第四条では、作成または取得から30年以内の「歴史的公文書」については、「氏名、生年月日その他の記述により特定の個人を識別することができるもの」の閲覧を制限している。さらに30年を経過したものについては、別表を設けて、秘密を公にすることによる当該個人の権利利益侵害の程度に応じて、閲覧制限期間を、①30年以上50年未満（イ. 国籍、人種または民族、ロ. 家族、親族または婚姻、ハ. 信仰、ニ. 思想、ホ. 伝染性の疾病、身体の障害その他の健康状態）、②50年以上80年未満（国籍、家族など）、③80年以上（イ. 門地、ロ. 遺伝性の疾病、精神の障害その他の健康状態、ハ. 犯罪歴または補導歴）の3段階に分けて規定していた。

2011年4月、『独立行政法人国立公文書館利用等規則』（平成23年4月1日規程第4号）が施行され、上で触れた『独立行政法人国立公文書館利用規則』（平成13年規程第7号）は廃止された。個人情報の公開に関しては、「第12条 二」に、『独立行政法人等の保有する情報の公開に関する法律』（平成13年法律第140号）で指定する項目にかかるもの以外の公開を認めている。個人に関する情報では、いくつかの除外項目があるものの、原則としては、「個人に関する情報（事業を営む個人の当該事業に関する情報を除く。）であって、当該情報に含まれる氏名、生年月日その他の記述等により特定の個人を識別することができるもの（他の情報と照合することにより、特定の個人を識別することができることとなるものを含む。）または特定の個人を識別することはできないが、公にすることにより、なお個人の権利利益を害するおそれがあるもの」と定め、従来のような細かい項目まで指定することはなくなっているようだ。

その後、情報通信技術の飛躍的な発展によってアクセスと分析が可能になった膨大な個人情報への利用期待が高まり、企業がビッグデータとして個人情報を利用できるよう、2015年、個人情報保護法が改正された。改正に並行して、個人情報の利用に関する考え方や世論も変化しつつある。高等教育においても、学生の教学データを統計的に分析して教育改善や経営判断に利用するなど、学生の個人情報を積極的に使う動きが出ている。学籍簿などの個票データに関しても、個人情報保護法の適用除外規定や独立行政法人等情報公開法を踏まえた上で、二次分析に耐えるデータベースの公開は可能になると考えられる。これについては、今後も国際的な動向⁶なども踏まえながらより適切な形を模索していくことが求められるだろう。

注

¹ 本研究は、兵庫県立神戸高等学校から史料の提供を受け、科学研究費「近代都市における中等教育機会の基礎的研究—個票データを用いた実業層の再生産戦略—」（基盤(C)課題番号20530788、研究代表者：中村隆文・平成22年度～24年度）の助成を得て行われたものである。データベース

の作成は、本研究メンバーに加え、神戸女子大学中村隆文研究室の大学院生・学生の協力を得て行われた。

² 自営（旧中間層）と会社員（新中間層）を区別するのが非常に難しいのは、自営の商店を株式会社化するケースにおいてである。たとえば白鶴酒造（神戸市東灘区）は1747（寛保3）年創業、1897（明治30）年に合名会社化し、1927（昭和2）年に株式会社化した。明治・大正期に白鶴酒造に勤務していた者は旧中間層に分類されるが、その後勤務を続け同じ仕事をしていても、昭和2年以降であれば会社員として新中間層に分類されることになる。このようなケースが神戸一中データでは非常に多く見られるのである。

³ 個票データに基づいた研究は土方（1994）が最初である。土方は、地方の小学校にのこる学籍簿と村の公式の役場資料とを対照させて学齢児童名簿を作成し、この村の学齢児童の悉皆調査を行った。中学校の学籍簿から個票データベースを作成して分析した研究は、鶴岡のこの研究が最初である。

⁴ 1995年から京都の国際日本文化研究センターを拠点におこなわれた共同研究で、日本の宗門改帳のように長期にわたって継続するすぐれた史料を持つ中国（東北部遼寧省）、イタリア（北部）、ベルギー（ワロン語地域）、スウェーデン（南部スカニア半島）の5つの社会の比較研究である。ユーラシア社会の人口・家族構造比較史研究（EAP）で行われた（黒須, 2008）。

⁵ 資料の詳細については別稿にゆずって詳しく記述することにしたい。

⁶ これに関してアメリカでは、「国民の税金を利用して遂行される研究はそのプロセスも成果も公開が原則である」が、「刑事犯罪データのようなトップシークレット情報」を扱う際、たとえば、社会科学データ共有の先駆的存在であるミシガン大学 ICPSR においては、数段階に分かれた公開・非公開の方法をとっている（黒須, 2008, p.54）。他方、「長年収集し整理されたデータを様々な人に大いに利用してもらいたいと願う一方、経験のない研究者がデータを間違えて利用したり解釈するかもしれない、というリスクを常に負うことになる」（Alter et al., 2006）という問題や、その際の責任の所在などについては未だ検討すべき課題として残っているとされる。

参考文献

Alter, George, Myron Gutmann, and Kees Mandemakers, 2006, “Problems and Possibilities for Distributing Longitudinal Historical Data.” Paper presented at HSN Workshop Disseminating and analyzing longitudinal historical data, Amsterdam, 21 March, 2006.

天野郁夫編（1991）『学歴主義の社会史』有信堂

五島敏芳（2007）「アーカイブズ情報の電子化・保存と共有化の動向」『情報知識学会誌』第17巻4号, pp.217-224.

速水融編著（2002）『近代移行期の家族と歴史』ミネルヴァ書房.

土方苑子（1994）『近代日本の学校と地域社会—村の子どもはどう生きたか—』東京大学出版会

土方苑子（2002）『東京の近代小学校 「国民」教育制度の成立過程』東京大学出版会.

広田照幸（1995）「教育・モダンティ・歴史分析」『教育社会学研究』第57集, pp.23-39.

広田照幸・鈴木智道・高瀬雅弘（2001）「旧制中学校卒業生の進路規定要因に関する研究」広田照幸編『近代化過程における中等教育の機能変容に関する地域間比較研究』pp.37-68.

ICPSR (Inter-University Consortium for Political and Social Research), 前田幸男訳（2007）「社会科学データ準備と保存のための手引き」東京大学社会科学研究所附属日本社会研究情報センター SSIJ データアーカイブ.

井上義和（2006）「旧制中学校進学機会における長男優先度の分析」『ソシオロジ』157号, pp.75-90.

- 加藤善子（2011）「近代日本における都市中学校生徒の社会的出自：旧制兵庫県立第一神戸中学校の学籍データによる分析」『信州大学人文社会科学研究』第5号, pp.175-189.
- 河野誠哉（2001）「鶴岡中学校在学者データベース—解題、ならびにその特質をめぐる基礎的検討—」広田照幸編『近代化過程における中等教育の機能変容に関する地域間比較研究』pp.169-1179.
- 菊地城司（2003）『近代日本の教育機会と社会階層』東京大学出版会.
- 黒須里美（2008）「長期マイクロデータをめぐる動向：歴史人口学研究の舞台裏」『人口学研究』第43号
- 望月厚志（1987）「近代日本における『教師層』の供給源—静岡県師範学校の事例をもとに—」『教育社会学研究』第42集, pp.215-229.
- 村越一哲（2007）「記録史料記述の標準化を促進するための提案」『国文学研究資料館紀要アーカイブズ研究篇』第3号, pp.61-73.
- 佐藤（粒来）香（2004）『社会移動の歴史社会学』東洋館出版社.
- 佐藤博樹・石田浩・池田謙一（2000）『社会調査の公開データ 2次分析への招待』東京大学出版会.
- 鈴江英一（2002）『近現代史料の管理と史料認識』北海道大学図書刊行会.
- 鈴木智道（2001）「個票型データのカテゴリー化をめぐる諸問題—『保護者職業』および『本人勤務先』—」広田照幸編『近代化過程における中等教育の機能変容に関する地域間比較研究』pp.191-215.
- 武石典史（2012）『近代東京の私立中学校—上京と立身出世の社会史』ミネルヴァ書房.
- 粒来香・佐藤俊樹（1995）「戦間期日本における職業と学歴」『教育社会学研究』第56集, pp.81-98.
- 寺崎里美・吉田文（2000）「落第と『半途退学』にみる旧制中学校の社会的機能」『教育社会学研究』第66集, pp.195-212.
- 保田卓・薄葉毅史・竹内洋（1999）「近代日本の学歴貴族の社会的出自と進路」『教育社会学研究』第65集, pp.49-67.
- 吉田文（1985）「明治～大正期の地域社会における中等教育の社会的機能」『教育社会学研究』第40集, pp.150-164.
- 湯田拓史（2010）『都市の学校設置過程の研究—阪神間文教地区の成立』同時代社.

ウェブページ

東京大学社会科学研究所附属日本社会研究情報センター『SSJ データアーカイブ（Social Science Japan Data Archive）』<http://jww.iss.u-tokyo.ac.jp/overview/2_2.html>

法令

『独立行政法人国立公文書館利用規則』平成13年4月2日規程第7号、最終改正平成18年6月20日.

『独立行政法人国立公文書館利用等規則』平成23年4月1日規程第4号、最終改正平成28年3月29日.

『独立行政法人等の保有する情報の公開に関する法律』平成13年12月5日法律第140号、

最終改正平成 28 年 11 月 28 日.

(加 藤 善 子 信州大学 総合人間科学系 高等教育研究センター 准教授)

(保 田 そ の 龍谷大学 非常勤講師)

2017 年 1 月 12 日受理 2017 年 2 月 14 日採録決定